The Computing Sector has created an overall IT Service Continuity Management Plan that covers the key areas that each individual plan would rely upon in a continuity situation such as command center information, vital records, personnel information. The purpose of this document is to describe the key information needed to recover this service in a business continuity situation once a decision to invoke has been made, and then to manage the business return to normal operation once the service disruption has been resolved.

Scope

Service Area: Scientific Data Storage and Access (SDSA).

Service Offerings: Enstore, dCache

- Enstore is tape based storage that can be accessed directly or through dCache
- dCache is a disk cache file system that can act as a front end file cache for Enstore or can be used as a non-tape backed file cache (volatile dCache)

Service Areas that depend on this service: Fermigrid (volatile dCache dependency)

dCache service instances provided by SDSA and covered in this plan are the Public and CDF dCache. The CMS dCache is not covered by this plan.

The SDSA Services are provided by the Data Movement and Development department (DMS), are operated by the DMS Storage Services Administration (SSA) group, and supported by the DMS Data Movement and Development (DMD) group, which develops Enstore and dCache software.

Recovery Objectives

Recovery Time Objective (RTO)

< 24 hours for critical service.

Recovery Point Objective (RPO)

< 24 hours.

As documented in the SLA (CS-doc-5032), users of the Scientific Data Storage and Access services are expected to be able to buffer 24 hours worth of data and to have the resources to catch up once service is restored.

Recovery Team

In this section describe the other services, roles, and responsibly required for recovering this service.

Service/Role/Function	Responsibility	Dependencies	Expected Response Time
Service Provider	Restore		< 24 hours
Service Provider Support Team	Restore hardware/OS/system software	DMS with OLA with FEF for dCache hardware and OS	< 4 hours
IT Server Hosting			
Network	Network connectivity, DNS	Cabling, network hardware	< 4 hours (see network SLA)
Facilities	Power and Cooling, equipment moves		< 4 hours?
Application Services	Authentication (Kerberos, Gums)	dCache, Enstore	< 4 hours. (HA failover – see authentication SLA)
External Service Provider	Restore functionality to the tape libraries, move equipment (tape drives)	Enstore tape libraries	< 4 hour
Managed Services	Eject, move and load tapes		< 8 hours?

The recovery team for the SDSA Services are:

1. Gene Oleynik (service owner) , Stan Naymola (service manager)
2. CS/SCD/SCF/DMS/SSA operations staff
3. CS/SCD/SCF/DMS/DMD development staff

Other teams that may need to stay in close contact are:

1. Tape library Vendors: Oracle
2. Managed Services
3. Network
4. Facilities

Recovery Strategy

Strategy for initial recovery

If the SDSA staff is the initial incident responder, contact will be made with the Service desk to declare the incident. If the incident is determined to be critical, a Critical Management Team will be formed. The SDSA responder to the incident will inform the Service Owner and work with the Management Team. The Management Team will decide whether this service recovery plan will be executed and will communicate that to the SDSA responder.

Overall recovery strategy

The SDSA responder forms a Service Recovery Team for the SDSA services. This team:

- Chooses a Recovery Coordinator to lead the team (nominally the SDSA staff initially responding to the incident).
- Performs a Damage Assessment (if not already done). Damage Assessment is described below.
- Sends notice of the outage to affected customers.
- Formulates a Recovery Plan from the damage assessment and the scenarios listed below.
    - Determines if alternate hardware or alternate data center is required.
    - Determines whether partial service restoration is required.
    - Notifies the Management Team of the plan.
- Works with Management Team to coordinate the recovery with other services.
- Takes steps and plan securing the environment in the tape library rooms if necessary (see Tape Environment Considerations below).
- Verifies underlying infrastructure is available. This includes facilities (power/cooling), networking (DNS, firewalls, routers, switches), and authentication services for recovery.
- Implements the recovery plan.
- Verifies environment and service readiness (as needed by the type and severity of the outage)
- Enables the service(s).
- Notifies the Management Team when the service(s) are available.
- If partial service was restored and alternate hardware deployed or alternate center used, when the hardware or center is repaired, restores services to normal operations according to the recovery plan.

Damage Assessment Procedure

In the case of a severe loss of service, the service recovery team is to investigate the problem, call in vendors if needed and utilize Computing Sector personnel expertise as required to assess the damage. The team follows this procedure:

1. Determine if there is potential for further disruption or damage to the service and take preventative steps if so.

2. Make a comprehensive list of the damage, damaged equipment, equipment that needs to be replaced, and any damage to tapes or their content, including potential further damage to tape due to degraded environmental conditions.
3. Make an assessment of the time to repair the affected portion of the SDSA service(s) and make it operational.
4. Make an assessment of the free resources in the SDSA services that are still operational and can be used to restore partial or full service to the affected portion.
5. When damage assessment has been completed, the Service Recovery Coordinator is to inform the Service Owner and Management Team of the assessment.

Communications

- If this Continuity Plan is to be activated, the Service Recovery Coordinator is to notify all team members and inform them of the details of the event.
- The Recovery Coordinator is to provide the Management Team with details of the Damage Assessment and resource availability and any immediate steps that need to be taken to prevent further damage.
- The Management Team in consultation with the other recovery teams will decide which resources to utilize for providing limited interim service. This process may engage customer representatives to make decisions about priorities and resources.
- Upon notification from the Service Recovery Coordinator, Team Leaders are to notify their respective teams as needed. Team members are to be informed of all applicable information and prepared to respond.
- The Service Recovery Coordinator is to notify remaining personnel (via notification procedures) on the general status of the incident.

High availability fail-over

The SDSA services do not provide HA fail-over

Recover at another site or multiple sites

Instances of the SDSA tape based services, Enstore, are located in two Fermilab data centers that are about 1 mile apart: The Feynman Computing Center second floor (FCC2) and the Grid Computing Center Tape Robot Room (GCC TRR). The services span and are active at both sites. If necessary, one of the sites can take on some of the functionality provided at the other. However, any significant recovery of functionality at the alternate site may necessitate the move of resources (servers, movers, tapes and tape drives). Restoring service to an alternate location is discussed in the Data Center Loss section below.

SDSA dCache services are all located in FCC2. New acquisitions may be located in the FCC3 data center.

Build from scratch

Most components of Enstore and dCache can be relocated or collocated on servers with other service components. In general, DMS/SSA attempts to bring the service back up on replacement hardware, which may be from on-site spares or a warranty replacement. In lieu of timely replacement hardware, the affected service component may be collocated to a server with other service components. This may degrade the service but will maintain continuity until replacement hardware can be deployed. In the case a system may need to be built from scratch, the steps are, for each affected system:

- Identify facilities resources where the equipment will go: rack, network and power
- Move and installation of all hardware if necessary
- Establish network connectivity for data LAN and private Control LANs
- Installation of the OS and system software (dCache:FEF, Enstore:DMS/SSA, ACSLS:Oracle)
- Establishing network connectivity for the service(s)
- Reconfiguration of cfengine and services if necessary.
- Reconfiguration of puppet for dCache OS and system software (by FEF) if necessary.
- Installation of the application(s) software from rpm server and ftp server
- Configuration of the system(s) through cfengine or puppet (FEF for dCache)
- Configuration of application(s) through cfengine/application configuration infrastructure
- Restoration of SDSA databases from backup (Enstore, Chimera, dCache)

This typically can be done within a several hours for a single server once hardware is available, powered and connected to the network. Building and testing a full service from scratch will, in general, take several to many days.

SDSA service software is distributed in the form of rpms. If rpms are not available, they can be built from the software code repository (currently CVS but soon to be moved to GIT) which is a service provided by the Scientific Computing Division. It is important to note that the software repository and distributions and their backups are maintained in the FCC2 data center - there are no alternate sites. Build from scratch is currently not possible if the repository or distributions in FCC2 are not available.

Recovery of RAID arrays may involve hardware replacement (controller/failed disks) or replacement of the entire array and restoration of the data from the failing hardware or from backups. In the latter case, due to the size and complexity of the databases, this may take a significant time for which service may be unavailable or degraded. Database restoration may also require involvement from the DMS Data Movement Development group (see Metadata Loss Considerations below).

Document sets CS-doc-5090 (Enstore) and CS-doc-5091 (dCache) contain further information on restoring these systems from scratch and metadata databases from backups.

Oracle maintains 3 ACSLS Sun/Oracle servers that are responsible for managing the tape libraries and with which Enstore communicates to direct tape mounts and dismounts and also provide for tape cartridge I/O operations. Each of these is a single point of failure for specific instances of Enstore. These

servers contain databases of tapes, drives and slots. These databases are backed up to a second local disk, and can also be recovered fairly quickly from the tape libraries themselves. These machines are under 24x7 maintenance for hardware, the ACSLS software, and operating system (Solaris). The process for building these from scratch is to contact Oracle via SNOW or directly off-hours to service the hardware.

Recovery Scenarios

Building not accessible (Data Center Available)

All service hosts are available remotely through console servers including remote power control. The tape libraries in the Data Centers may need physical intervention.

- Contact the Crisis Center. The Computing Sector Continuity plan lists the location of these centers as well as alternate locations.
- Execute the overall strategy for recovery.
- Restore the tape libraries and other hardware that may require physical intervention when the building becomes accessible.

Data Center Failure (Building Accessible)

- Contact the Crisis Center. The Computing Sector Continuity plan lists the location of these centers and alternates.
- Execute the overall strategy for recovery using the Data Center Failure procedure.

Building not accessible and Data Center Failure

- Contact the Crisis Center. The Computing Sector Continuity plan lists the location of these centers and alternates.
- Execute the overall strategy for recovery using the Data Center Failure procedure
- Restore the tape libraries and other hardware that may require physical intervention when the building becomes accessible

Critical recovery team not available

- If initial respondent to the incident, contact the service desk or Emergency Services
- Wait for the Critical Recovery team to set up a Crisis Management Team which will determine whether this Continuity plan should be executed
- Execute the rest of the overall strategy for recovery according for the appropriate scenario if called on to do so

Most of the work can be done remotely. Server and mover power can be cycled remotely, but some may require on-site presence (possibly for escorting tape libraries vendors, for example).

Pandemic (during normal operations)

The SDSA services require specialized knowledge to operate. Most operational tasks can be completed remotely through VPN access. Some operational tasks require the presence of staff on site, in particular tape library operations and tape cartridge I/O operations.

- Vendor service personnel require an escort into the tape library rooms
- Flipping/un-flipping cartridge write protect tabs requires physical presence
- Entering blank tapes requires physical presence

If necessary, these activities could be managed by unskilled staff under the remote direction of skilled staff.

Metadata Loss Considerations

Backups of the metadata (Chimera/pnfs namespace and Enstore databases) for each instance of the SDSA services are made daily. Recent backups are available on disk on the instance of the SDSA service and are copied to tape daily to both GCC and FCC tape libraries. Restores from backup are exercised frequently as part of managing the service. The restoration process is to

1. Restore the hardware if needed and perform tests to make sure it is sound.
2. Restore the OS and system software.
3. Restore the service application software.
4. Follow the procedure to restore the database from backup.
5. Validate the restore database is functional.
6. Plan and perform post-restoration work, if needed, to backfill entries in the database.

Tape Environment Considerations

Data are written to tape on tracks that are less than 5 microns in width and pitch. The readability of a track is sensitive to the dimensional stability of the media. Tape media dimensions will vary with changes in humidity and temperature. A tape drive may have difficulty reading a tape at a humidity and temperature significantly different than what it was written at. This can result in poor read performance or read failure.

Recommendations from the tape and tape library manufacturers are:

1. Operating Ideal 72°F, 45% RH. Recommended Range: 68-77°F, 40-50%RH
2. Maximum ranges: 60-90°F, 20-80% RH
3. Recommended acclimation period: 72 hours

In the case of a loss of climate control, steps should be taken to keep the climate in the tape library rooms within the maximum ranges listed above, and as close as possible to the recommended ranges. Depending on the ambient outside temperature and humidity and the length of the outage, the tapes may need to be acclimated to the environment for up to 72 hours before they can be used. Tapes should

not be written to until they are acclimated. Exceptions may be necessary if writing is urgent and the risk is deemed acceptable.

The high-density tapes used in these libraries are delicate. T10000T2 tapes, for example, are 5.2 microns thick and the track pitch is 3.5 micron. Subjecting a tape to shock by dropping it more than a few feet can damage the edge of the tape and make the data on it unreadable or unusable for writes. Bulk moves of tapes should be avoided where possible, but if necessary, the tapes should be handled with care and transported in cases designed specifically for the transport of data tapes.

Tape Library not operational

If tape libraries at either FCC or GCC data center are not operational or the data center is inaccessible for a prolonged period of time, partial service may be restored to the other center (or other alternate libraries in the same data center if they are viable).

If a tape library were not operational, it would be a difficult task to remove specific tapes, including blanks. Enstore does not know where tapes are located in the tape library; only the tape library management interface database (e.g. ACSLS database) knows the mapping between a tape's label and the slot in the tape library where it resides. Once identified, the tapes would need to be manually located and removed by physically entering the library. In addition, moving tape does have associated risk (see Tape Environment Considerations section). Relocating tapes would not be feasible for a temporary relocation, but could be for a prolonged period such as if the data center or tape library is permanently damaged.

Most data has a single copy at one of these data centers (CMS, Run II). Single copy data is in general either statistical in nature and in large enough quantity that the loss of single tape is not statistically significant, or are backups from another site (for both external and internal customers, e.g. DESDM and local RMAN backups of scientific databases). Other data is duplicated between the data centers. Data that may be duplicated includes:

- Statistical data for which loss of single tape is statistically significant (e.g. MiniBooNe tankdata)
- Small Datasets (e.g. BeamsTools)
- Data that is not statistical and for which file loss is significant

Data that falls into these categories are duplicated (or are being duplicated) to both data centers. This can be done automatically by Enstore, or manually by the experiment (e.g. MINOS makes their own copies). In general, the primary copy that the user accesses is located at FCC and the secondary copy at GCC.

For files that are duplicated by Enstore, in the event of the loss or inaccessibility of primary copies, the secondary copies could be made available. This requires intervention by the administrators.

The high level strategy for this Tape Library not operational scenario is:

1. Pause the affected Enstore Library Managers.
2. Contact the vendor to move any tape drive resources, as decided by the management and recovery teams, to libraries into the alternate libraries.
3. Arrange with Managed Services to move any blanks or tapes to be read from the failed library to the alternate library(s). This step may only be feasible for a more permanent reallocation of tapes (see above). Instead, blank tapes for writes may need to be rush ordered.
4. If primary copies of Enstore duplicated files exist in the affected library and secondary copies exist in an operational library:
    a. Make a list of these primary copy files and store it in backed up storage (for future restoration of normal operations)
    b. Administratively swap the roles of the primary and secondary roles in the metadata so that users can access the files in the operational library
5. Install additional mover computers for any additional tape drives that have been added to the alternate library.
6. Reconfigure the affected Enstore instances' logical tape library managers to point to the media changers in the alternate tape libraries.
7. Pause the alternate Enstore logical library managers
8. Configure the tape drive movers of drives in the alternate tape library so that a portion of the tape drives are dedicated for use by affected logical Enstore libraries. The partitioning of drives is based on the resource allocation agreed upon by the management team.
9. Allocate blank and relocated tapes to these logical libraries
10. Resume the affected library managers

Roles of other teams

Managed services team, with the assistance of the Service Owner (in selecting, locating and removing tapes):

- Moves blank tapes from the failed tape libraries to alternate tape libraries for write access recovery as necessary. Load rush ordered blanks.
- Moves selected subsets of written tapes from the failed libraries to the alternative libraries for a read-access recovery as necessary.

Networking Team:

- Replaces and repairs any damaged network elements required for the contingency
- Configures the network for any moved or new computers required to establish the contingency plan as necessary
- Reconfigures the private control LAN topology as necessary.

Tape Library Vendor:

- Moves tape drives from the failed to the alternative library(s) as necessary

Data Center Loss

The storage services are located in the second floor of FCC2 and in the GCC TRR. These two centers are located about 1 mile apart on the Fermilab site. The SDSA services span and are active at both of these sites. Scientific Data is stored in tape libraries located at both of these sites. Most of the SDSA service's server computers are located on the second floor of FCC, while mover computers that transfer data are collocated with the tape libraries at both locations. The locations of SDSA equipment are:

- FCC2
  - Enstore servers and metadata raid arrays. There are 6 or more servers for each instance of enstore and there are 3 instances of enstore. Most of these servers perform critical functions for the service and have critical state data on the RAID arrays.
  - dCache servers. These servers perform critical functions and are single points of failure.
  - dCache pool nodes (disk file servers) and associated RAID
  - 30,000 slots capacity in a cluster of 3 tape libraries (Public and Run II non-RAW data)
  - Movers for tape drives in the FCC tape libraries
  - Infrastructure nodes for managing the service (migration, scanning, etc.)
  - ACSLS (Oracle tape management server) for the tape libraries in FCC
- GCC TRR
  - 40,000 slot capacity in a cluster of 3 tape libraries (CMS) and one stand-alone tape library (Public duplicate secondary copies and backups and Run II RAW data)
  - Movers for the tape drives in the GCC tape libraries
  - Two ACSLS machines – one for CMS, one for everything else (Run II RAW, Public backups, scientific database RMAN backups)
  - A small number of infrastructure nodes, Enstore test system server computers

Moving the services hosted at one of these centers to the other may result in additional load and service degradation, but some level of service continuity could be maintained. Hardware may need to be moved between the sites in order to provide adequate resources or to establish service(s). In addition, except where noted below, data on tapes at the failed center would not be accessible unless the tapes were moved to the alternative center. The procedure for Tape Library not operational would be followed.

If the GCC data center or the Tape Robot Room located there were inaccessible, Enstore and dCache services can continue to function degraded. Tape storage capacity in GCC would be unavailable and the data on the tapes contained in the libraries at GCC (CMS, Run II RAW, Public backups, and scientific database RMAN backups) would be inaccessible. The procedure for the Tape Library not operational would be followed.

If the FCC second floor data center were unavailable, the Enstore and dCache services would also be unavailable since they are all located in FCC2. These services are provided by many computers (3 instances x 6+ servers + RAID arrays for Enstore and at least a dozen for dCache). For these reasons, if the FCC2 data center were unavailable, the SDSA services would be unavailable until the data center was

restored. If FCC2 were lost permanently or for a very long period, building the SDSA services from scratch may make sense and the Build from scratch procedure would be followed. In this case, in addition to following the Build from scratch and Tape library not operational procedures, limited resources would need to be prioritized:

1. Form a team to assess the impact and which will work with customers to form a disaster recovery plan to return partial essential service and to eventually return full service
2. Document the impact to the customers and the resources that are available to restore partial service
3. Meet with the customers to come up with a mutually satisfactory plan to share resources and bring up partial essential service
4. Acquire additional resources if necessary (e.g. blank tapes)
5. Formulate a plan to restore full service and meet with customers to get agreement on the plan

Return to Operations

Document any requirements and tasks that would need to be completed in order to return to operations. If you have procedures for returning to operations after a continuity situation occurs, then you can reference them here.

1. If this recovery is from a partial degraded interim recovery to a full recovery, the following additional steps must be made to restore full service
    a. Restore any relocated service components and server computers back to their original location and configuration
    b. Replace any mover computers, tape drives for the repaired libraries if any were moved during the recovery
    c. Restore any modified duplicated file's secondary and primary roles for files that had this change made for partial recovery
    d. Restore the affected Enstore instances mover configurations as necessary.
    e. Schedule moving any relocated tapes back to the recovered libraries with the Managed Services team or on an emergency basis.
2. Establish all dependent services are operational and running (LAN, Private LAN, authentication and facilities)
3. Rebuilt file systems must fsck OK.
4. Monitoring must indicate that service components are fully functional or that there is acceptable partial functionality (e.g. some monitoring) in some components. In general enstore servers and dCache admin nodes should be fully operational.
5. > 50% of Enstore mover computers for a given library are available.
6. Tape libraries are all up and inventoried (exceptions may be made for partial restoration)
7. Test transfers of files (Enstore and/or dCache) complete successfully.

8. Tape acclimation time has transpired
9. Resume the repaired instance

Tape library vendor

- Return tape drives from the alternate library to the recovered library if necessary

Managed Services Team

- Move the blanks, written tapes, and any other tapes that were originally moved in the recovery from the alternate library to the recovered libraries as directed by request from the Service Recovery Team.

Document Change Log

Version	Date	Author(s)	Change Summary
1.1	7	7/8/2013	Minor post review changes